Requested Patent: JP9282105A

Title:

DISK STORAGE SYSTEM AND RECOVERY METHOD FOR ERROR CORRECTION DATA PREPARATION PROCESSING THEREFOR ;

Abstracted Patent: JP9282105 ;

Publication Date: 1997-10-31 ;

Inventor(s):

KITAMURA MANABU; YAMAMOTO AKIRA; YAMAMOTO YASUTOMO; SATO TAKAO ;

Applicant(s): HITACHI LTD ;

Application Number: JP19960086086 19960409 ;

Priority Number(s): ;

IPC Classification: G06F3/06; G06F3/06; G06F3/06; G06F12/16 ;

Equivalents: ;

ABSTRACT:

PROBLEM TO BE SOLVED: To guarantee data in the case of occurren of a fault during preparing a parity in a storage system for tentatively storing updating data sent from a processor in a cache memory, preparing a new parity and then storing the updating data and the parity in a storage device.SOLUTION: A nonvolatile cache 33 is provided with respective areas for storing the data after updating and before updating and the parities after updating and before updating. A parity preparation control part 38 prepares the parity after updating from the data before and after updating and the parity before updating, then turns ON a nonvolatile identifier for indicating parity preparation completion and copies the new parity and the data after updating to respective areas before updating. After the processing of a storage controller 3 is interrupted, a recovery control part 39 controls whether to re-execute a parity preparation processing from the start or re-execute it from the processing after turning ON the identifier by the OFF or ON of the identifier.

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平9－282105

(43)公開日　平成９年(1997)10月31日

(51)Int.Cl.⁶		識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ	3/06	３０５		Ｇ０６Ｆ	3/06	３０５Ｃ
		３０２				３０２Ａ
		５４０				５４０
	12/16	３２０	7623－5B		12/16	３２０Ｌ

審査請求　未請求　請求項の数９　ＯＬ　（全 15 頁）

(21)出願番号　特願平8－86086

(22)出願日　平成８年(1996)４月９日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　北村　学
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

(72)発明者　山本　彰
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

(72)発明者　山本　康友
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

(74)代理人　弁理士　薄田　利幸

最終頁に続く

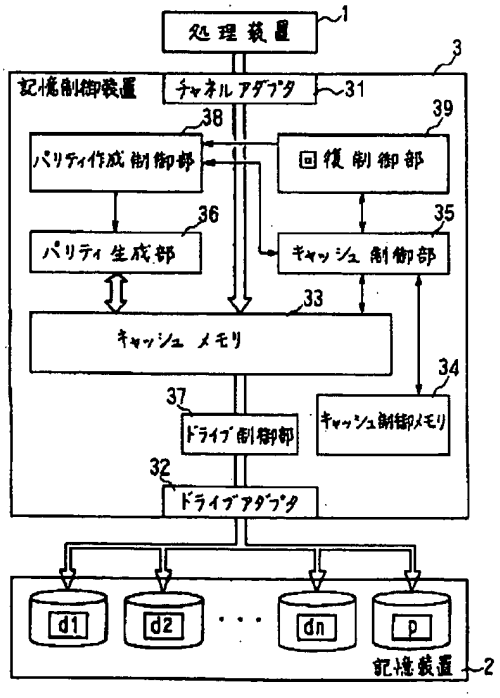
(54)【発明の名称】　ディスク記憶システム及びその誤り訂正データ作成処理の回復方法

(57)【要約】

【課題】　処理装置１から送られた更新データを一旦キャッシュメモリ３３に格納し、新しいパリティを作成してから更新データとパリティとを記憶装置２に格納する記憶システムにおいて、パリティ作成中に障害が発生した場合のデータの保証をする。

【解決手段】　不揮発のキャッシュ３３は、更新後及び更新前のデータ、更新後及び更新前のパリティを格納するそれぞれの領域を有する。パリティ作成制御部３８は、更新前後のデータ及び更新前のパリティから更新後のパリティを作成した後、パリティ作成完了を示す不揮発の識別子をオンにして新しいパリティと更新後データをそれぞれの更新前領域にコピーする。記憶制御装置３の処理が中断した後、回復制御部３９は識別子のオフ又はオンによってパリティ作成処理を始めから再実行するか又は識別子をオンにした後の処理から再実行するかを制御する。

図１



【特許請求の範囲】

【請求項１】データ及び誤り訂正データを格納するディスク記憶装置と、処理装置と該ディスク記憶装置との間に介在し、該処理装置と該ディスク記憶装置との間のデータ転送を制御する記憶制御装置とを有するディスク記憶システムであって、該記憶制御装置は、該ディスク記憶装置に反映していない更新データを格納する第１の領域と、該更新データに対応し該ディスク記憶装置に反映済みのデータを格納する第２の領域と、該更新データに対応する誤り訂正データを格納する第３の領域と、該反映済みのデータに対応する誤り訂正データを格納する第４の領域とを有する不揮発性のキャッシュメモリと、第１の領域に格納される更新データ、第２の領域に格納される反映済みのデータ及び第４の領域の誤り訂正データから更新後の誤り訂正データを作成して第３の領域に格納する第１の処理ステップと、第３の領域の更新後の誤り訂正データを第４の領域へ上書きし、第１の領域の更新データを第２の領域へ上書きする第２の処理ステップとを実行する処理手段とを有するディスク記憶システムにおいて、

該記憶制御装置は、さらに第１の処理ステップが終了したか否かの状態を記憶する不揮発性の識別子と、第１の処理ステップが終了したとき該識別子を終了状態に設定する処理手段と、障害によって該記憶制御装置の処理が中断された後、該識別子を参照して終了状態でなければ該処理手段の第１の処理ステップを再実行し、終了状態であれば該処理手段の第２の処理ステップを再実行するよう制御する回復処理手段とを有するディスク記憶システム。

【請求項２】第３の領域及び第４の領域の誤り訂正データに対応して第１の領域及び第２の領域が複数個存在し、各第２の領域及び第４の領域の内容がそれぞれディスクアレイを構成する別々のディスク記憶装置に格納されることを特徴とする請求項１記載のディスク記憶システム。

【請求項３】処理装置とデータ及び誤り訂正データを格納するディスク記憶装置との間に介在し、該処理装置と該ディスク記憶装置との間のデータ転送を制御する記憶制御装置であって、該記憶装置に反映していない更新データを格納する第１の領域と、該更新データに対応し該ディスク記憶装置に反映済みのデータを格納する第２の領域と、該更新データに対応する誤り訂正データを格納する第３の領域と、該反映済みのデータに対応する誤り訂正データを格納する第４の領域とを有する不揮発性のキャッシュメモリと、第１の領域に格納される更新データ、第２の領域に格納される反映済みのデータ及び第４の領域の誤り訂正データから更新後の誤り訂正データを作成して第３の領域に格納する第１の処理ステップと、第３の領域の更新後の誤り訂正データを第４の領域へ上書きし、第１の領域の更新データを第２の領域へ上書きする第２の処理ステップとを実行する処理手段とを有する記憶制御装置において、

該記憶制御装置は、さらに第１の処理ステップが終了したか否かの状態を記憶する不揮発性の識別子と、第１の処理ステップが終了したとき該識別子を終了状態に設定する処理手段と、障害によって該記憶制御装置の処理が中断された後、該識別子を参照して終了状態でなければ該処理手段の第１の処理ステップを再実行し、終了状態であれば該処理手段の第２の処理ステップを再実行するよう制御する回復処理手段とを有する記憶制御装置。

【請求項４】該記憶装置は、複数のディスク記憶装置から構成されるディスクアレイであり、第３の領域及び第４の領域の誤り訂正データに対応して第１の領域及び第２の領域が複数個存在し、各第２の領域及び第４の領域の内容が別々のディスク記憶装置に格納されることを特徴とする請求項３記載の記憶制御装置。

【請求項５】該キャッシュメモリは、第１の領域、第２の領域、第３の領域及び第４の領域を有する第１のキャッシュメモリと、第１のキャッシュメモリと同じ内容の情報を格納する第２のキャッシュメモリとから構成され、該処理手段は、第１のキャッシュメモリと第２のキャッシュメモリのうち正常なキャッシュメモリについて第１の処理ステップ及び第２の処理ステップを実行し、該回復処理手段は、障害によって該記憶制御装置の処理が中断された後、第１のキャッシュメモリと第２のキャッシュメモリのうち正常なキャッシュメモリについて再実行するよう制御することを特徴とする請求項１記載のディスク記憶システム。

【請求項６】該キャッシュメモリは、第１の領域、第２の領域、第３の領域及び第４の領域を有する第１のキャッシュメモリと、第１のキャッシュメモリと同じ内容の情報を格納する第２のキャッシュメモリとから構成され、該処理手段は、第１のキャッシュメモリと第２のキャッシュメモリのうち正常なキャッシュメモリについて第１の処理ステップ及び第２の処理ステップを実行し、該回復処理手段は、障害によって該記憶制御装置の処理が中断された後、第１のキャッシュメモリと第２のキャッシュメモリのうち正常なキャッシュメモリについて再実行するよう制御することを特徴とする請求項３記載の記憶制御装置。

【請求項７】データ及び誤り訂正データを格納するディスク記憶装置と、処理装置と該ディスク記憶装置との間に介在し、該処理装置と該ディスク記憶装置との間のデータ転送を制御する記憶制御装置とを有するディスク記憶システムであって、該記憶制御装置は、該ディスク記憶装置に反映していない更新データを格納する第１の領域と、該更新データに対応し該ディスク記憶装置に反映済みのデータを格納する第２の領域と、該更新データに対応する誤り訂正データを格納する第３の領域と、該反映済みのデータに対応する誤り訂正データを格納する第

４の領域とを有する不揮発性のキャッシュメモリを含むディスク記憶システムの誤り訂正データ作成方法において、
第１の領域に格納される更新データ、第２の領域に格納される反映済みのデータ及び第４の領域の誤り訂正データから更新後の誤り訂正データを作成して第３の領域に格納する第１のステップと、
第１のステップが終了したときその状態を示す識別子を不揮発性の記憶装置に格納する第２のステップと、
第３の領域の更新後の誤り訂正データを第４の領域へ上書きし、第１の領域の更新データを第２の領域へ上書きする第３のステップと、
障害によって該記憶制御装置の処理が中断された後、該識別子を参照して終了状態でなければ第１のステップを再実行し、終了状態であれば第３のステップを再実行する第４のステップとを有することを特徴とするディスク記憶システムの誤り訂正データ作成処理の回復方法。
【請求項８】第３の領域及び第４の領域の誤り訂正データに対応して第１の領域及び第２の領域が複数個存在し、各第２の領域及び第４の領域の内容がそれぞれディスクアレイを構成する別々のディスク記憶装置に格納されることを特徴とする請求項７記載のディスク記憶システムの誤り訂正データ作成処理の回復方法。
【請求項９】該キャッシュメモリは、第１の領域、第２の領域、第３の領域及び第４の領域を有する第１のキャッシュメモリと、第１のキャッシュメモリと同じ内容の情報を格納する第２のキャッシュメモリとから構成され、第１のステップから第４のステップまでの処理は、第１のキャッシュメモリと第２のキャッシュメモリのうち正常なキャッシュメモリについて行うことを特徴とする請求項７記載のディスク記憶システムの誤り訂正データ作成処理の回復方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、ディスク記憶システム及びその誤り訂正データ作成処理に係わり、特に誤り訂正データ作成処理の回復方法に関する。
【０００２】
【従来の技術】従来、デビット・Ａ・パターソンらは、高速に大量のデータを複数のディスクにアクセスするディスク装置のディスク故障時におけるデータの冗長性を実現するディスクアレイ装置についての論文「低価格ディスクの冗長アレイ（ＲＡＩＤ）の一事例（Ａ Ｃａｓｅ ｆｏｒ Ｒｅｄｕｎｄａｎｔ Ａｒｒａｙｓ ｏｆ Ｉｎｅｘｐｅｎｓｉｖｅ Ｄｉｓｋｓ（ＲＡＩＤ））」を１９８８年６月１日から３日にシカゴで開催されたＡＣＭ ＳＩＧＭＯＤで発表している。提案されたディスクアレイ装置は、ＲＡＩＤ（Ｒｅｄｕｎｄａｎｔ Ａｒｒａｙｓ ｏｆＩｎｅｘｐｅｎｓｉｖｅ Ｄｉｓｋｓ）と略称される。ＲＡＩＤの仕組みを簡単に説明すると次のようなものである。
【０００３】ＲＡＩＤでは、ライトするデータを複数のデータに分割し、この分割した複数のデータについて排他的論理和をとることによってパリティを計算する。そして分割したデータとパリティを並列に複数のディスクに対して書き込む。ライトデータをｎ個に分割した場合、それらのデータをｄ１，ｄ２，．．．，ｄｎとするとこの分割したデータからパリティｐを生成するには次の（１）式に従って行い、パリティｐの作成後に、ｄ１，ｄ２，．．．，ｄｎとｐはそれぞれ別のディスクに書き込まれる。
【０００４】

$$p = d1 (+) d2 (+) \ldots (+) dn \qquad (1)$$

ただし（＋）は、排他的論理和を表す。排他的論理和は、各項を２進数に変換した時の各桁をそれぞれ参照し、１が奇数個の時は１、偶数個の時は０になる。例えばデータを３分割にして書き込む場合、ｄ１＝０１，ｄ２＝１１，ｄ３＝００の時(それぞれ２進数)にはｐは１０になる。このｄ１，ｄ２，．．．，ｄｎとｐのデータの集合をパリティグループといい、ｐは常に（１）式に従った値になっていなければならない。そのため例えば既に書き込みの行われた領域ｄ１のみに別のデータＤが上書きされた場合でも、パリティｐについて新規パリティＰを算出してＤとともにＰもディスクに書き込む必要がある。その場合は次の（２）式によりパリティを算出してＤとＰをディスクに書き込む操作をする。
【０００５】

$$P = D (+) d1 (+) p \qquad (2)$$

特開平４－２４５３５２号公報には、記憶装置であるディスク装置がＲＡＩＤ構成をとりキャッシュメモリを有する記憶制御装置のライト処理の方法が記載されている。キャッシュメモリを有する記憶制御装置は、中央処理装置からのライト指示に対し記憶制御装置内のキャッシュメモリにライトデータを格納した時点で中央処理装置に対してライト処理終了を報告する。そして記憶制御装置は、中央処理装置からのライト指示とは非同期にディスク装置へライトデータを格納する。この制御をライトアフタ制御といい、ディスク装置へのデータ転送を行う前に中央処理装置に対してライト処理終了を報告するため、中央処理装置への高速な応答が可能になる。
【０００６】この記憶制御装置は、パリティを作成するのに必要十分なデータをキャッシュメモリ上に格納した後パリティを作成する。以下ライト処理について説明する。ディスクに書き込み済みのデータｄ１の一部又は全部に対して書き込み要求が来てデータｄ１に対する更新データＤがキャッシュメモリ上に書き込まれると、記憶制御装置は中央制御装置へライト完了報告を行い、その後に非同期にパリティの作成を行う。記憶制御装置は上

式（２）に従ってパリティを作成するため、旧データｄ１及び旧パリティｐの置かれる領域と新たに作成されるパリティＰを置く領域をキャッシュメモリ上に確保して、ディスクからｄ１及びｐを読み込む。それから（２）式に従い新規にパリティＰを作成して確保したキャッシュメモリ上の領域に格納する。その後、更新データＤと新規作成パリティＰをキャッシュメモリ上のｄ１及びｐのある領域へ上書きする処理を行ってから、ＤとＰをディスク装置へ書き込む。このときｄ１とｐのデータが既にキャッシュメモリ上にある場合には、新たにキャッシュを確保する処理とｄ１とｐをディスクから読み込む処理が省略できるために処理が高速化される利点がある。

【０００７】

【発明が解決しようとする課題】ＲＡＩＤ構成の記憶装置では、前述の通りデータの一部を書き換えるとともにそれに対応するパリティグループのパリティデータも常に前述の（１）式に従っていなければならないので、データの更新と同時に書き換える必要がある。ただし実際にはデータの書き換えとパリティの書き換えは完全に同時に実行できるわけではなく、一時的にはパリティだけ又はデータだけが先にディスクに書き込まれているという状態が存在する。これは記憶制御装置がキャッシュメモリを有する場合にキャッシュメモリの内容についても同じことが言える。ライトデータＤからパリティＰを作成してｄ１とｐの置かれているキャッシュメモリの領域にコピーする時、Ｄだけがコピーされた時点で停電等による処理中断が発生した場合、パリティ作成前にはｄ１とｐが置かれていた領域はＤとｐが存在するという状態になる。そのために例えばこの状態から同じパリティグループ内の別のデータｄ２に対する更新データＤ２からのパリティＰ２の作成を行うと、本来は

Ｐ２＝ｄ２（＋）Ｄ２（＋）Ｐ

を行わなければならないのに、Ｐの更新前パリティデータｐを使った

ｄ２（＋）Ｄ２（＋）ｐ

を行って間違ったパリティを作成してしまうため、このようなパリティ作成途中での停電等による処理中断や障害発生でのデータ損失の問題は十分考慮される必要がある。しかしながら従来技術では、パリティ作成中の停電等による処理中断が発生した場合、特に更新データＤとＤに対するパリティデータＰを、更新前データｄ１とｄ１に対するパリティｐのおかれている領域にコピーしている際のデータ損失やその際の回復処理について考慮されていない。

【０００８】このｄ１やｐがディスク装置に書き込み済みの状態の場合、ｄ１及びｐの置かれているキャッシュメモリにディスク装置からその内容を再度読み込んで、パリティの作成処理を始めからやり直しをすれば良いことは容易に考えられるので考慮する必要はない。しかしライトアフタ制御ではキャッシュメモリにライトデータを格納した時点で中央処理装置に対してライト処理終了を報告するため、ライトデータＤからパリティＰを作成して、それらがｄ１やｐのある領域にコピーされた直後に同じ領域にさらに新しいライトデータＤ’がキャッシュメモリ上に書き込まれてそれに対する新規パリティＰ’を作成する処理が始まることもあり得る。その時に上の例のようにＤ’をＤのある領域にコピーした直後で障害が発生した場合にパリティ作成をやり直そうとしても、ＤとＰはディスク上には置かれていないデータなので読み込むことはできず、やり直しはできない。

【０００９】本発明の目的は、このようなディスク装置への書き込みが行われていないデータに対する更新データからパリティ作成を行っている途中に障害が発生した場合のデータ保証をすることにある。

【００１０】

【課題を解決するための手段】本発明は、不揮発のキャッシュメモリ上に置かれた更新データＤ、ディスク記憶装置に反映済みの対応するデータｄ及びｄに対応する誤り訂正データｐから更新後の誤り訂正データＰを作成する第１のステップと、第１のステップが終了したときその状態を示す識別子を不揮発性の記憶装置に格納する第２のステップと、誤り訂正データＰをｐの領域へ上書きし、更新データＤをｄの領域へ上書きする第３のステップと、障害によって記憶制御装置の処理が中断された後、この識別子を参照して終了状態でなければ第１のステップを再実行し、終了状態であれば第３のステップを再実行する第４のステップとを有することを特徴とするディスク記憶システムの誤り訂正データ作成処理の回復方法を特徴とする。

【００１１】本発明によれば、誤り訂正データの作成が終了したか否かを境界にして回復処理が再実行を開始するポイントを変えており、いずれの場合にもデータ及びパリティの正しいことが保証され、再開始後の処理においてもディスク記憶システムは間違ったパリティを作成することはない。

【００１２】なお同一パリティグループ内に更新データがＤ１，Ｄ２，・・・のように複数ある場合には、第１のステップでは更新データＤ１，Ｄ２，・・・、それぞれ反映済みの対応するデータｄ１，ｄ２，・・・及び誤り訂正データｐから更新後の誤り訂正データＰを作成し、第３のステップでは更新データＤ１，Ｄ２，・・・をそれぞれｄ１，ｄ２，・・・へ上書きし、その他の処理は上記の通りである。

【００１３】

【発明の実施の形態】以下、本発明の実施形態について図面を用いて説明する。

【００１４】（１）第１の実施形態

図１は、第１の実施形態のディスク記憶システムの構成図である。ディスク記憶システムは、記憶制御装置３及

び記憶装置2から構成される。処理装置1は、記憶装置2に対してリード／ライト要求を発行する情報処理装置である。記憶装置2は、複数のディスク記憶装置を有し、ディスクアレイを構成する。記憶制御装置3は、処理装置1及び記憶装置2に接続され、チャネルアダプタ31、ドライブアダプタ32、キャッシュメモリ33、キャッシュ制御メモリ34、キャッシュ制御部35、パリティ生成部36、ドライブ制御部37、パリティ作成制御部38及び回復制御部39から構成される。図1で太い矢印はデータの流れを示し、細い矢印は制御情報の流れを示す。ただし図面を簡単にするために本発明と関連の少ない制御の流れについて省略している。チャネルアダプタ31は、チャネルインタフェースによって処理装置1のチャネルに接続され、処理装置1と情報の送受信を行う。ドライブアダプタ32は、ドライブインタフェースによって記憶装置2に接続され、記憶装置2と情報の送受信を行う。キャッシュメモリ33は、バッテリにより不揮発化された記憶装置であり、記憶装置2に書き込むデータ及びその誤り訂正データを一時的に格納する。以下誤り訂正データをパリティと呼ぶ。キャッシュ制御メモリ34は、不揮発化された記憶装置であり、キャッシュメモリ33上のデータ及びパリティを管理する情報を格納する記憶装置である。キャッシュ制御部35は、キャッシュメモリ33上の記憶領域を管理し、キャッシュ制御メモリ34上に管理情報を作成し、更新する。パリティ生成部36は、データ更新に伴ってキャッシュメモリ33上のデータ及びパリティから新しいパリティを生成する。ドライブ制御部37は、記憶装置2内のディスク記憶装置とキャッシュメモリ33との間のデータの転送を制御する。パリティ作成制御部38は、パリティ作成処理を制御し、そのステータスを管理する。回復制御部39は、停電等の障害が発生してパリティの作成が中断したとき、パリティ作成処理を回復する。

【0015】ここで以下の説明に使用する用語の説明をする。レコードとは、処理装置1と記憶制御装置3との間で転送するデータの単位である。スロットとは、キャッシュメモリ33上の記憶領域の単位であるとともに、処理装置1から来たライトデータを記憶装置2の各ディスク記憶装置に分割する際の大きさの単位である。スロットのサイズはレコードのサイズより大きく、スロットは複数のレコードを収容可能である。すべてのスロットは同じ大きさである。各スロットにはユニークなスロット番号が付与される。処理装置1からリード／ライトされるデータが置かれるスロットをデータスロットといい、このデータスロットから作成されるパリティが置かれるスロットをパリティスロットという。スロットはさらに細かいブロックに分割される。ブロックはディスク記憶装置と記憶制御装置3との間でデータを転送するときの最小単位である。同一スロット内の各ブロックにはブロック番号が付与される。スロット番号とブロック番号が与えられると、特定のディスク記憶装置上のデータ格納場所が決定される。1つのパリティを作成するときに対象となる複数スロットに亘るデータの集合をそのパリティとともにまとめてパリティグループという。キャッシュメモリ33上のデータのうち記憶装置2上のデータと一致しているデータをクリーンデータという。記憶装置2に反映されておらず、キャッシュメモリ33上にのみ記憶されている更新データをダーティデータという。ダーティデータのうちパリティの作成されているデータを物理ダーティデータ、パリティの作成されていないデータをホストダーティデータという。

【0016】処理装置1から1レコード又は複数レコードから成るデータのライト要求が来たとき、記憶制御装置3はこのデータの大きさが1スロットに収まる範囲の大きさであればそのまま、複数スロットにまたがる大きさの場合にはスロット境界でデータｄ1，ｄ2，・・・，ｄnに分割し、一旦キャッシュメモリ33に記憶する。分割したデータはそれぞれキャッシュメモリ33上の連続するスロット番号をもつ別々のスロットに格納される。1つのレコードはスロット内の1つ又は複数の連続するブロックを占有する。キャッシュメモリ33へのデータ書き込みが終了した時点で記憶制御装置3は、処理装置1にライト終了を報告し、その後処理装置1のライト指示とは非同期に単一のデータスロット又は同一のパリティグループに属する複数のデータスロットからパリティを作成する。パリティ作成制御部38はパリティ生成部36にパリティの生成を要求する。パリティ生成部36は、キャッシュメモリ33上のデータを読み出してパリティを作成し、キャッシュメモリ33上のパリティスロットにこのパリティを書き込む。パリティの作成が終了した後、ドライブ制御部37は各データスロット内のデータとパリティスロット内のパリティを各々別々のディスク記憶装置に格納する。

【0017】図2は、キャッシュメモリ33内に確保されるスロットの構成とキャッシュ制御メモリ34内に確保されるスロット管理情報の構成を示す図である。各スロット42は、それぞれ分割したデータ及びパリティを含む。1つのスロット42は、ライト面44とリード面45から構成される。ライト面44は処理装置1からのライト要求によって受け取ったデータを格納する領域、リード面45はディスク記憶装置に書き込むデータを格納する領域である。ライト面44はホストダーティデータを格納し、リード面45は物理ダーティデータ又はクリーンデータを格納する。ライト面44又はリード面45は、さらに複数のブロック46に分割されている。分割したデータ及びパリティは、それぞれライト面44及びリード面45内の1つ又は複数のブロックを占有する。

【0018】キャッシュ制御メモリ34内に確保されるスロット管理情報43は、スロット番号、ライト面ポイ

ンタ、リード面ポインタ及びスロットステータス４７から構成され、それぞれ対応するスロットに置かれる分割されたデータ又はパリティを管理する。スロット番号は管理するスロットの番号である。ライト面ポインタは、ライト面４４内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納する。またリード面ポインタは、リード面４５内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納する。スロットステータス４７は、スロット内で管理しているデータの状態を示す情報を保有しており、チャネルライト中、ディスクリード中、ディスクライト中、パリティ作成中、パリティ作成完了、ホストダーティ及び物理ダーティを示す識別子から構成される。各識別子は、その状態がオンかオフかの２値をもつ。チャネルライト中、ディスクリード中、ディスクライト中は、それぞれ処理装置１からキャッシュメモリ３３へのデータ転送中、記憶装置２からキャッシュメモリ３３へのデータ転送中、キャッシュメモリ３３から記憶装置２へのデータ転送中の状態を示す。

【００１９】処理装置１からライト要求が来ると、記憶制御装置３は指定されたシリンダ番号／トラック番号／レコード番号のようなディスク・アドレスをスロット番号／ブロック番号に変換し、処理装置１から受け取った１つ又は複数のレコードをそのまま又はｎ個のデータに分割する。次にキャッシュ制御部３５はキャッシュ制御メモリ３４上にこのスロット番号から始まるｎ個のスロット管理情報４３を作成する。この操作をスロット確保という。次にキャッシュ制御部３５は、キャッシュメモリ３３の空いている領域からｎ個のスロット４２のライト面４４を確保し、各ライト面４４上でレコードが開始するブロックのアドレスとブロック数をスロット管理情報４３のライト面ポインタに格納する。この操作をキャッシュ確保という。キャッシュ確保をした後に記憶制御装置３は、各スロット４２上のブロックに分割したデータの各々を書き込む。次にパリティ作成制御部３８は、キャッシュ制御部３５を介してパリティスロットとライト面、リード面のキャッシュを確保し、パリティ生成部３６はパリティを作成又は更新してパリティスロットのライト面に書き込む。次にパリティ作成制御部３８はパリティスロットのライト面４４のパリティをリード面４５にコピーし、各データスロットのライト面４４のデータをリード面４５にコピーする。この後スロット管理情報４３のライト面ポインタの内容をクリアして確保したキャッシュをポインタから外す。この操作をキャッシュ解放という。次にデータスロットのリード面４５上のデータ及びこのパリティグループに属するパリティスロットのリード面４５上のパリティが記憶装置２に書き込まれ、スロット管理情報４３のリード面ポインタの内容をクリアして各スロットのリード面のキャッシュを解放する。一方処理装置１からリード要求が来ると、スロット４２のリード面のキャッシュ確保をして記憶装置２からデータ及びパリティを読み込み、リードデータが不要になるとそのリード面のキャッシュを解放する。

【００２０】次にパリティ作成制御部３８及びパリティ生成部３６が行うパリティ作成処理について説明する。パリティ作成処理は、大別するとフェーズ１とフェーズ２の２つの過程に分けられる。フェーズ１は、パリティ作成の対象となるデータスロットに対応してキャッシュメモリ３３上にパリティスロットを確保し、データスロットの内容からパリティを作成する。フェーズ２は、データスロットのライト面のホストダーティデータとパリティスロットのライト面に作成したパリティを各々のスロットのリード面にコピーして物理ダーティデータとし、各スロットのライト面のキャッシュを解放する処理を行う。なお以下の説明を簡単にするために、１つのデータスロット内のレコードを更新する場合を主体として説明する。

【００２１】図３及び図４は、フェーズ１の処理の流れを示すフローチャートである。パリティ作成制御部３８は、キャッシュ制御部３５を介して対象となるパリティグループについてパリティスロットのスロット管理情報４３を作成してパリティスロットを確保し、そのスロット管理情報４３のスロットステータス４７のうち「パリティ作成中」と「ホストダーティ」をオンにする（ステップ１０１）。「パリティ作成中」は、スロットが他の処理によって操作されることを禁止するために用い、「ホストダーティ」状態は回復処理を行う時にこの状態を参照して処理を決定するために用いられる情報である。次にそのパリティスロットのリード面とライト面のキャッシュメモリが確保されているか否かをチェックし、確保されていない場合にはキャッシュ制御部３５を介して確保する（ステップ１０２）。次に同一パリティグループ中のデータスロットについて、そのデータスロットに対応するスロット管理情報４３のスロットステータス４７が「ホストダーティ」であるデータスロットを選択し、選択したデータスロットに対応するスロットステータス４７のうちの「パリティ作成中」をオンにし（ステップ１０３）、そのデータスロットのリード面とライト面のキャッシュメモリが確保されているか否かをチェックし、確保されていない場合にはキャッシュ制御部３５を介して確保する（ステップ１０４）。次にパリティ作成対象のデータスロットに対応するスロット管理情報４３のライト面ポインタを参照し、ホストダーティデータの存在する範囲を算出する（ステップ１０５）。すなわちデータスロットのライト面でレコードが開始するブロックのアドレスとブロック数を得る。次にステップ１０５で算出した範囲に対応するリード面に物理ダーティ又はクリーンデータがあるかどうかを判定し、存在しない場合には（ステップ１０６Ｎｏ）、記憶装置２からクリーンデータをリード面に読み込む（ステップ１０

7）。次にステップ105で算出したデータスロットのライト面のデータの存在する範囲について、対応するパリティスロットのリード面にパリティが存在するか確認する（ステップ108）。存在しない場合には（ステップ108No）、記憶装置2からその範囲のクリーンデータをパリティスロットのリード面に読み込む（ステップ109）。次にパリティ作成制御部38はパリティ生成部36に要求してパリティを作成する（ステップ110）。パリティ生成部36は、データスロットのリード面、データスロットのライト面及びパリティスロットのリード面のデータを読み出して指定された範囲のパリティを作成し、新しいパリティをパリティスロットのライト面に書き込む。

【0022】図5及び図6は、フェーズ2の処理の流れを示すフローチャートである。ステップ110の処理が終了するとすぐにパリティスロットのスロット管理情報43のスロットステータス47の「パリティ作成完了」をオンにする（ステップ201）。スロット42のライト面へのデータ出力が終了した直後にパリティスロットのライト面に作成された新しいパリティをそのリード面の対応する一連のブロックへ上書き（コピー）する（ステップ202）。次にデータスロットのホストダーティデータをそのリード面の対応する一連のブロックへ上書き（コピー）する（ステップ203）。次にパリティスロットについて不要となったライト面の一連のブロックをスロットから解放し（ステップ204）、パリティスロットのスロット管理情報43のスロットステータス47の「ホストダーティ」をオフ、代わりに「物理ダーティ」をオンする（ステップ205）。次にデータスロットについて不要となったライト面の一連のブロックをスロットから解放し（ステップ206）、スロットステータス47の「ホストダーティ」をオフにし、「物理ダーティ」をオンにする（ステップ207）。最後にパリティスロットとデータスロットのスロットステータス47の「パリティ作成中」をオフにして（ステップ208）、パリティ作成処理を終了する。

【0023】なおパリティグループ内の1つのスロット内のレコードを更新する場合に、キャッシュ上の同一パリティグループ内の他のデータスロットに「ホストダーティ」のまま残っているレコードがあれば、ステップ103～104の処理をすべてのデータスロットについて繰り返し、またステップ105～107の処理をすべてのデータスロットについて繰り返し、ステップ108～110はデータスロット内ですべてのホストダーティのデータが存在する範囲についてパリティを作成する必要がある。またステップ203，206及び207もパリティ作成の対象となるすべてのデータスロットについて処理を行う必要がある。

【0024】次にパリティ作成処理中の障害発生により処理が中断された後のスロット回復処理について説明する。記憶制御装置3に障害が発生すると、すべての処理は中断されて不揮発化されているキャッシュメモリ33とキャッシュ制御メモリ34以外は初期状態に戻る。記憶制御装置3の動作が初期状態から開始されると、まず回復処理が起動する。

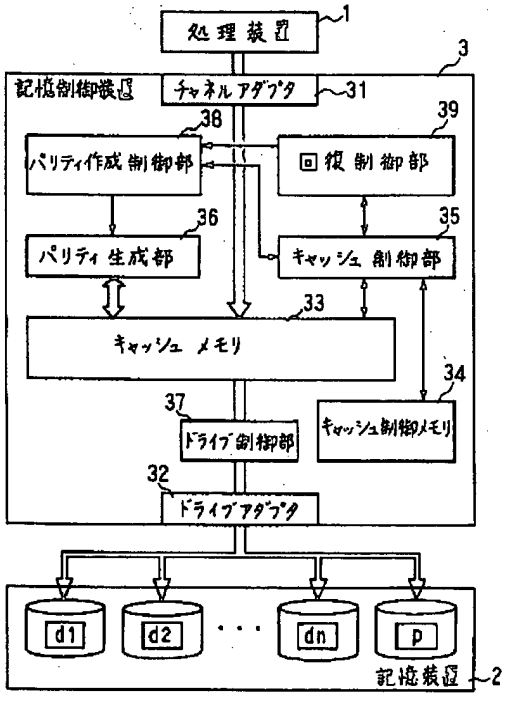
【0025】図7は、回復制御部39の処理の流れを示すフローチャートである。回復制御部39は、スロットステータス47が「パリティ作成中」状態のパリティスロットと、そのパリティグループ内のデータスロットでスロットステータス47の「パリティ作成中」がオンになっているすべてのスロットを取り出す（ステップ1001）。次にパリティスロットのスロットステータス47の「パリティ作成完了」がオンでなければ（ステップ1002No）、対象とするスロットステータス47をパリティ作成前の状態に戻す（ステップ1003）。すなわち「パリティ作成中」がオン状態のデータスロットについてはこれをオフにし、パリティスロットについては「パリティ作成中」をオフにし、「ホストダーティ」もオフにすることによってパリティ作成前の状態に戻る。そしてパリティ作成処理をフェーズ1の最初から再実行すればよい（ステップ1004）。すなわちパリティ作成処理がフェーズ1の途中で終っていた場合、パリティ作成制御部38はリード面のデータに操作を行っていないので、再びパリティ作成処理を始めから行ってもデータの整合性が保たれており問題ない。一方パリティスロットのスロットステータス47の「パリティ作成完了」がオンであれば（ステップ1002Yes）、パリティスロットのスロットステータス47の「ホストダーティ」がオンであれば（ステップ1005Yes）、B以降（ステップ202以降）を実行する（ステップ1006）。パリティスロットのスロットステータス47の「ホストダーティ」がオフであれば（ステップ1005No）、C以降（ステップ205以降）を実行する（ステップ1007）。パリティスロットのスロットステータス47の「パリティ作成完了」がオンであればフェーズ1は完了しているから、フェーズ2だけを実行してスロットをパリティ作成完了の状態にすればデータの整合性は保たれる。ここでパリティスロットのスロットステータス47の「ホストダーティ」がオフであれば、ステップ204までの処理は済んでいるのであるから、C以降を再実行すればよい。またパリティスロットのスロットステータス47の「ホストダーティ」がオンであれば、ステップ205の処理まで達していないので、B以降を再実行する必要がある。なおパリティスロットのスロットステータス47の「パリティ作成完了」がオンであるときには、ステップ202のパリティのリード面へのコピーが終了している可能性があるから、フェーズ1の最初から再実行すると、以降のパリティ作成に際して間違ったパリティ作成を行う可能性があることが理解される。

【００２６】（２）第２の実施形態
以下、第１の実施形態をベースとしてキャッシュメモリが二重化されている場合のパリティ作成制御部３８及び回復制御部３９の処理について述べる。キャッシュが二重化されている場合には、パリティの作成中に一方のキャッシュメモリに障害が発生して使用できないとき他方のキャッシュメモリの内容を使って回復を行い、処理続行することが可能である。以下第１の実施形態に対応する構成要素がある場合には同じ参照番号を用いる。
【００２７】図８は、第２の実施形態のディスク記憶システムの構成図である。キャッシュメモリ３３－１及びキャッシュメモリ３３－２は、二重化されたキャッシュメモリである。以下キャッシュメモリ３３－１及びキャッシュメモリ３３－２をそれぞれキャッシュメモリ１及びキャッシュメモリ２又はキャッシュ１及びキャッシュ２と略称することがある。キャッシュ制御部３５に含まれるキャッシュ障害判定部３５１は、キャッシュメモリ３３－１及びキャッシュメモリ３３－２が正常か否かを判定し、片方のキャッシュに障害があって使用できないときには、正常な方のキャッシュを選択する。
【００２８】図９は、キャッシュ制御メモリ３４内に確保されるスロット管理情報の構成及びキャッシュメモリ３３内のスロットとの関連を示す図である。スロット管理情報４３は、キャッシュメモリ１上のスロットとキャッシュメモリ２上のスロットの両方を管理する。すなわちキャッシュ１のライト面ポインタはキャッシュ１上のライト面４４内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納し、リード面ポインタはキャッシュ１上のリード面４５内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納する。またキャッシュ２のライト面ポインタはキャッシュ２上のライト面４４内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納し、リード面ポインタはキャッシュ２上のリード面４５内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納する。スロット番号及びスロットステータス４７については第１の実施形態で説明した通りである。
【００２９】処理装置１からライト要求が来ると、キャッシュ制御部３５はキャッシュメモリ３３－１及びキャッシュメモリ３３－２のそれぞれについて１個又はｎ個のスロット４２のライト面４４を確保し、各ライト面４４上のブロック・アドレスとブロック数をそのスロットに対応するスロット管理情報４３のキャッシュ１のライト面ポインタ及びキャッシュ２のライト面ポインタに格納してキャッシュを確保する。その後、記憶制御装置３は同一のデータをキャッシュ１のライト面とキャッシュ２のライト面とに重複して書き込む。
【００３０】図１０及び図１１は、キャッシュが二重化されており、キャッシュ１とキャッシュ２がともに正常であるときのパリティ作成制御部３８及びパリティ生成部３６が行うパリティ作成処理のフェーズ１の処理の流れを示すフローチャートである。第１の実施形態と異なる点は、キャッシュを確保するときに、キャッシュ１、キャッシュ２の順に確保し（ステップ４０２及び４０４）、キャッシュにデータを読み込むときにもキャッシュ１、キャッシュ２の両方に読み込みする（ステップ４０７及び４０９）点である。またパリティ生成部３６はキャッシュ１とキャッシュ２の両方についてそれぞれパリティを作成する（ステップ４１０）。
【００３１】図１２及び図１３は、図１０及び図１１に続くフェーズ２の処理の流れを示すフローチャートである。第１の実施形態と異なる点は、ライト面のデータをリード面にコピーするとき、キャッシュ１とキャッシュ２の両方について行い（ステップ５０２及び５０３）、キャッシュのライト面を解放するとき、キャッシュ１とキャッシュ２の両方について行う（ステップ５０４及び５０６）点である。
【００３２】図１４は、第２の実施形態の回復制御部３９の処理の概略を示すフローチャートである。回復制御部３９は、キャッシュが障害状態にあるかどうかチェックする（ステップ２００１）。両方とも正常ならば（ステップ２００１Ｙｅｓ）、図１５に示すキャッシュが二重化されているときの回復処理を行う。キャッシュ１のみが正常ならば（ステップ２００２Ｙｅｓ）、キャッシュ１について回復処理を行う。キャッシュ２のみが正常ならば（ステップ２００２Ｎｏ）、キャッシュ２について回復処理を行う。キャッシュ１又はキャッシュ２についての回復処理は、それぞれのキャッシュについて第１の実施形態の回復処理（図７）と同じである。
【００３３】図１５は、キャッシュが二重化されているときの回復制御部３９の回復処理の流れを示すフローチャートである。回復処理は基本的に第１の実施形態と同じであるが、キャッシュ１とキャッシュ２の両方について回復処理を行う。
【００３４】第２の実施形態によれば、キャッシュ１、キャッシュ２のうち１つに異常が発生していて使用できないときには、キャッシュ障害判定部３５１は、正常な方のキャッシュを選択するので、パリティ作成制御部３８は選択された正常のキャッシュについてパリティ作成処理を続行できる。記憶制御装置３に障害が発生して処理が中断されたとき、回復制御部３９はこの正常な方のキャッシュについて回復処理を行う。
【００３５】
【発明の効果】本発明によれば、誤り訂正データの作成が終了したか否かを境界にしてそれぞれ適切なポイントからパリティ作成処理を再実行するので、正しいデータ及びパリティが保証される。
【図面の簡単な説明】
【図１】第１の実施形態のディスク記憶システムの構成図である。

【図２】第１の実施形態のキャッシュメモリ３３及びキャッシュ制御メモリ３４の構成を示す図である。
【図３】第１の実施形態のパリティ作成処理（フェーズ１）の処理の流れを示すフローチャートである。
【図４】第１の実施形態のパリティ作成処理（フェーズ１）の処理の流れを示すフローチャート（続き）である。
【図５】第１の実施形態のパリティ作成処理（フェーズ２）の処理の流れを示すフローチャートである。
【図６】第１の実施形態のパリティ作成処理（フェーズ２）の処理の流れを示すフローチャート（続き）である。
【図７】第１の実施形態の回復制御部３９の処理の流れを示すフローチャートである。
【図８】第２の実施形態のディスク記憶システムの構成図である。
【図９】第２の実施形態のキャッシュメモリ３３及びキャッシュ制御メモリ３４の構成を示す図である。
【図１０】第２の実施形態のパリティ作成処理（フェーズ１）の処理の流れを示すフローチャートである。
【図１１】第２の実施形態のパリティ作成処理（フェーズ１）の処理の流れを示すフローチャート（続き）である。
【図１２】第２の実施形態のパリティ作成処理（フェーズ２）の処理の流れを示すフローチャートである。
【図１３】第２の実施形態のパリティ作成処理（フェーズ２）の処理の流れを示すフローチャート（続き）である。
【図１４】第２の実施形態の回復制御部３９の処理の流れを示すフローチャートである。
【図１５】第２の実施形態の回復制御部３９の処理の流れを示すフローチャート（続き）である。
【符号の説明】
３：記憶制御装置、３３：キャッシュメモリ、３４：キャッシュ制御メモリ、３６：パリティ生成部、３８：パリティ作成制御部、３９：回復制御部、４２：スロット、４３：スロット管理情報、４４：ライト面、４５：リード面、４７：スロットステータス

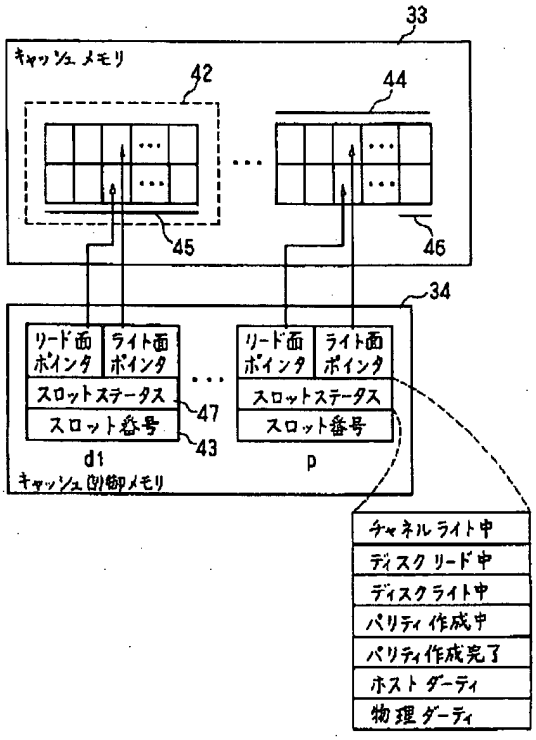
【図１】

図１



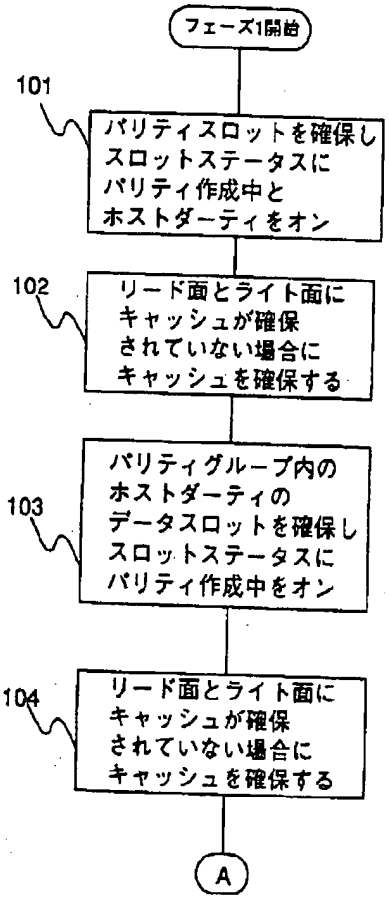
【図２】

図２



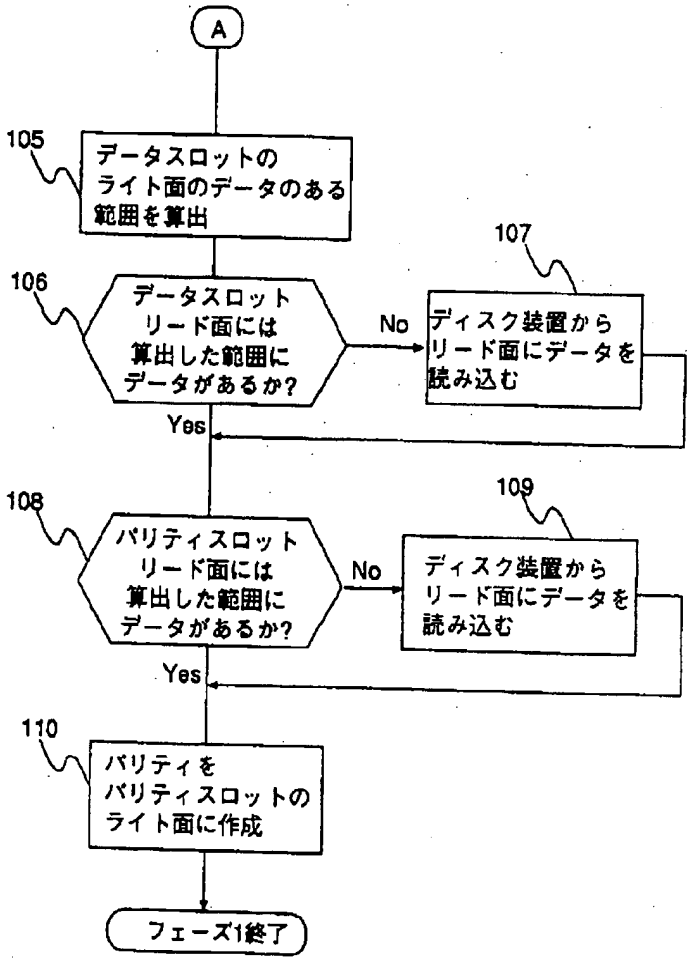
【図3】

図3


フェーズ1開始
101
パリティスロットを確保し
スロットステータスに
パリティ作成中と
ホストダーティをオン
102
リード面とライト面に
キャッシュが確保
されていない場合に
キャッシュを確保する
103
パリティグループ内の
ホストダーティの
データスロットを確保し
スロットステータスに
パリティ作成中をオン
104
リード面とライト面に
キャッシュが確保
されていない場合に
キャッシュを確保する
A

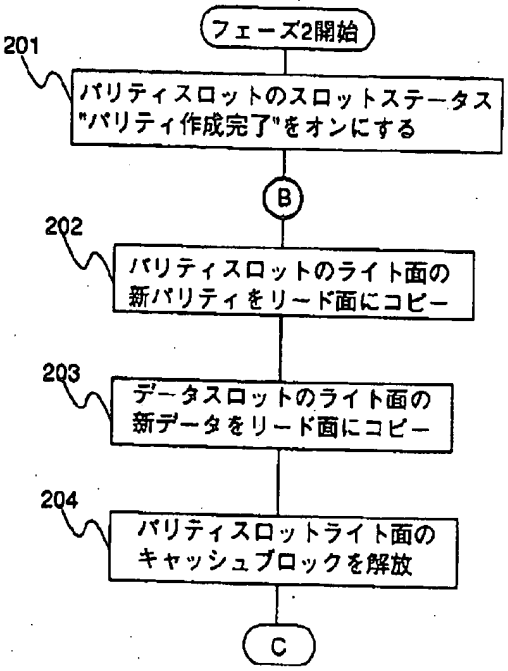
【図4】

図4


A
105
データスロットの
ライト面のデータのある
範囲を算出
106
データスロット
リード面には
算出した範囲に
データがあるか?
No
107
ディスク装置から
リード面にデータを
読み込む
Yes
108
パリティスロット
リード面には
算出した範囲に
データがあるか?
No
109
ディスク装置から
リード面にデータを
読み込む
Yes
110
パリティを
パリティスロットの
ライト面に作成
フェーズ1終了

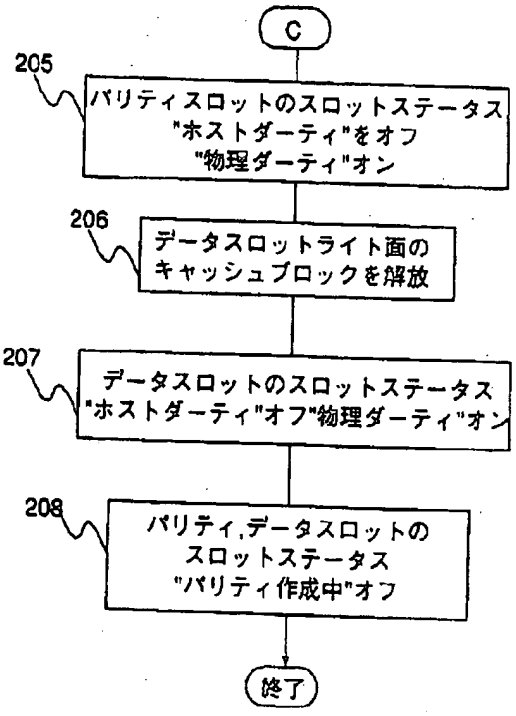
【図5】

図5



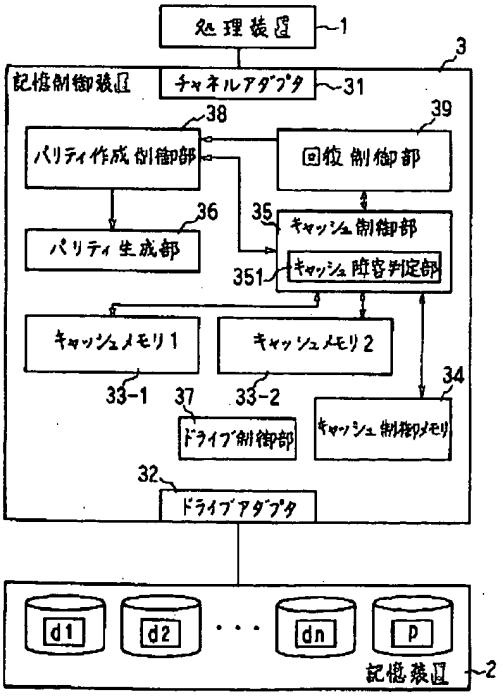
【図6】

図6



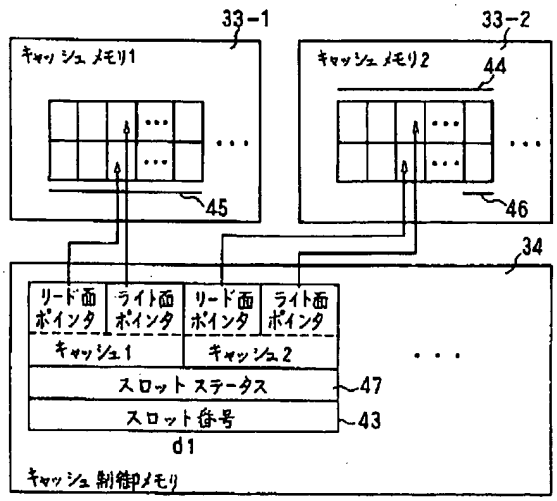
【図8】

図8



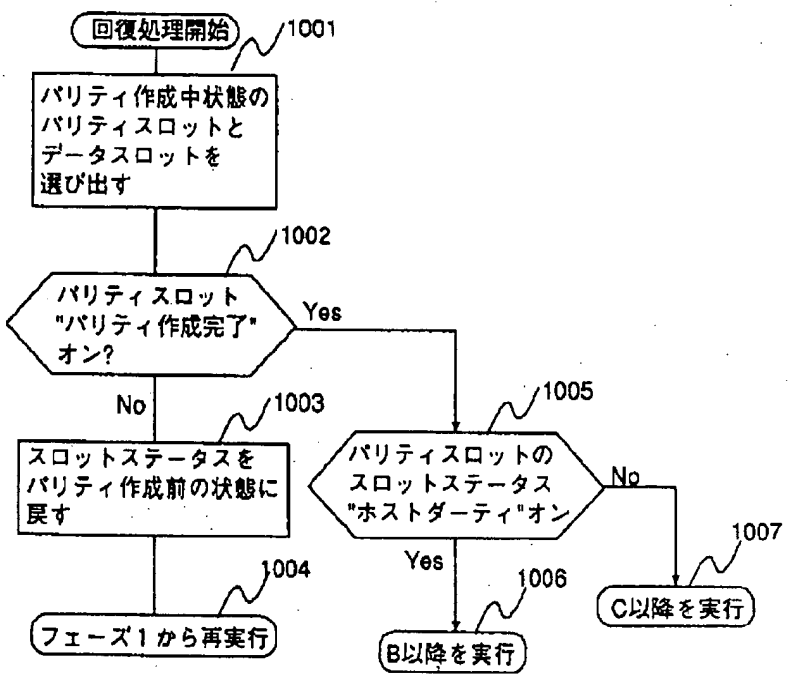
【図9】

図9



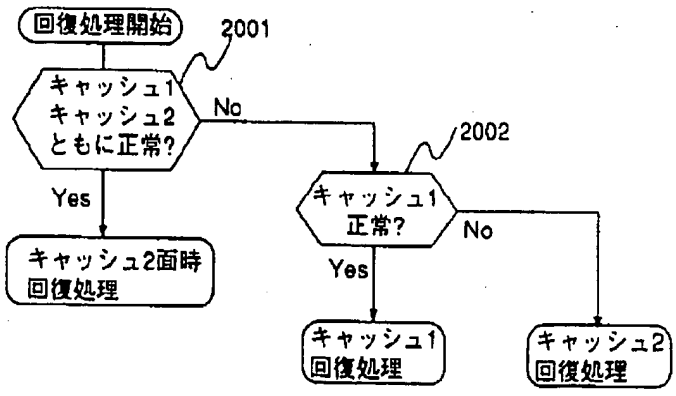
【図7】

図7


回復処理開始
1001
パリティ作成中状態の
パリティスロットと
データスロットを
選び出す
1002
パリティスロット
"パリティ作成完了"
オン?
Yes
No
1003
スロットステータスを
パリティ作成前の状態に
戻す
1004
フェーズ1から再実行
1005
パリティスロットの
スロットステータス
"ホストダーティ"オン
No
Yes
1006
B以降を実行
1007
C以降を実行

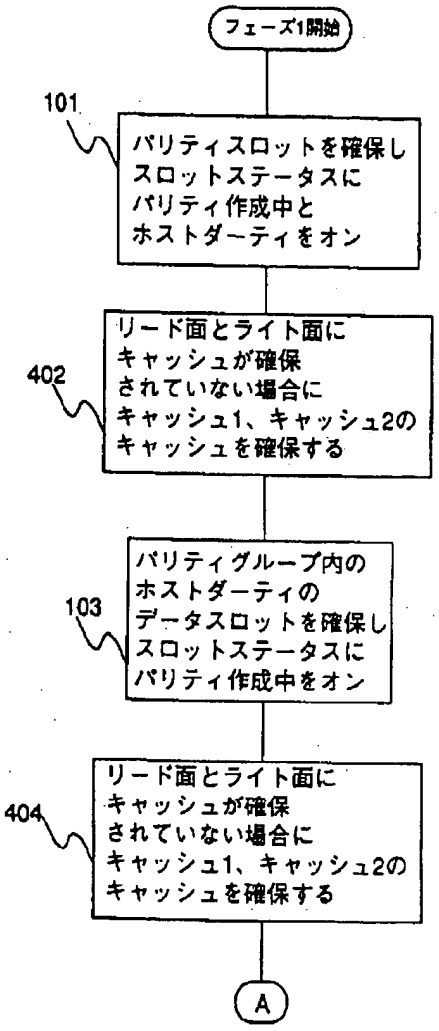
【図14】

図14


回復処理開始
2001
キャッシュ1
キャッシュ2
ともに正常?
No
Yes
キャッシュ2面時
回復処理
2002
キャッシュ1
正常?
No
Yes
キャッシュ1
回復処理
キャッシュ2
回復処理

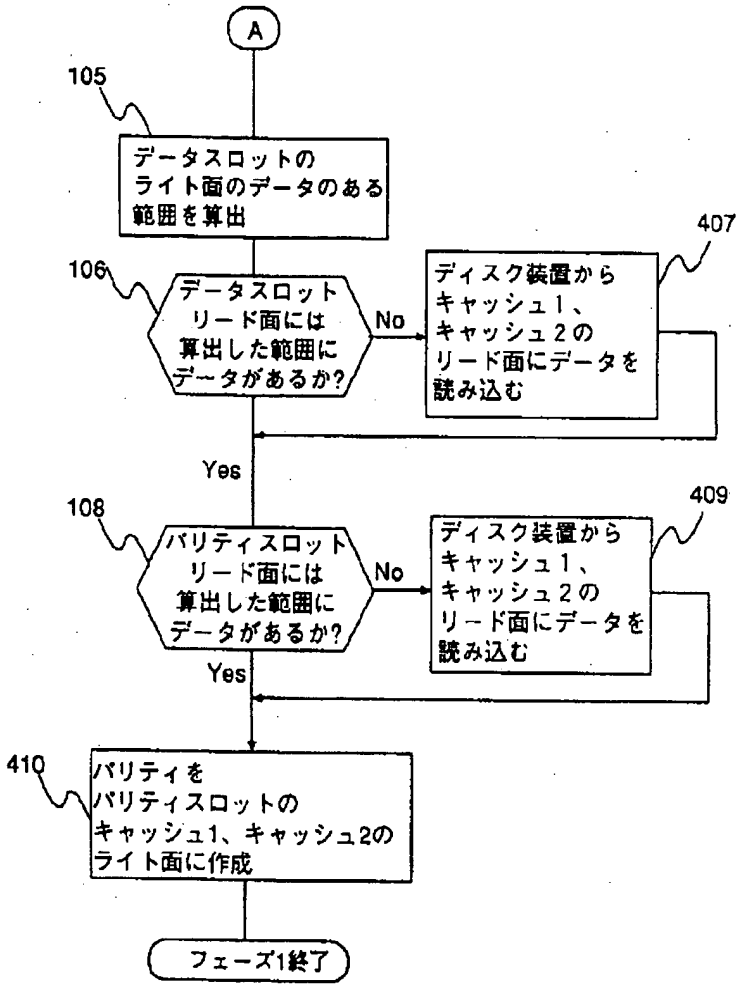
【図10】

図10



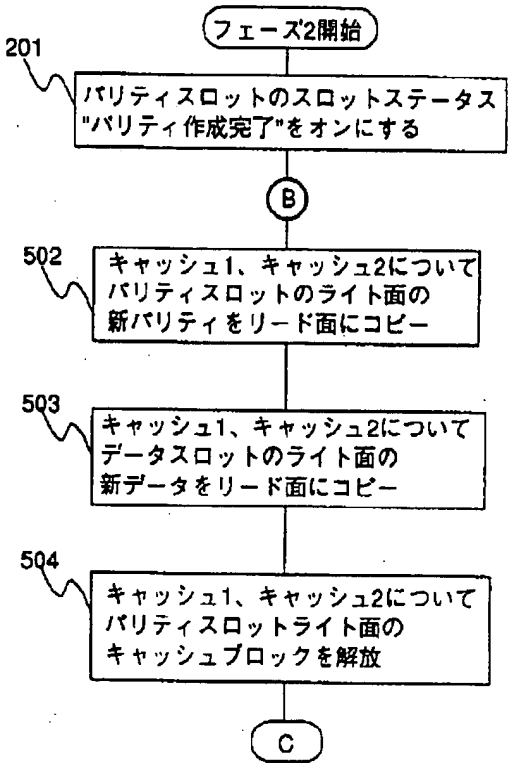
【図11】

図11



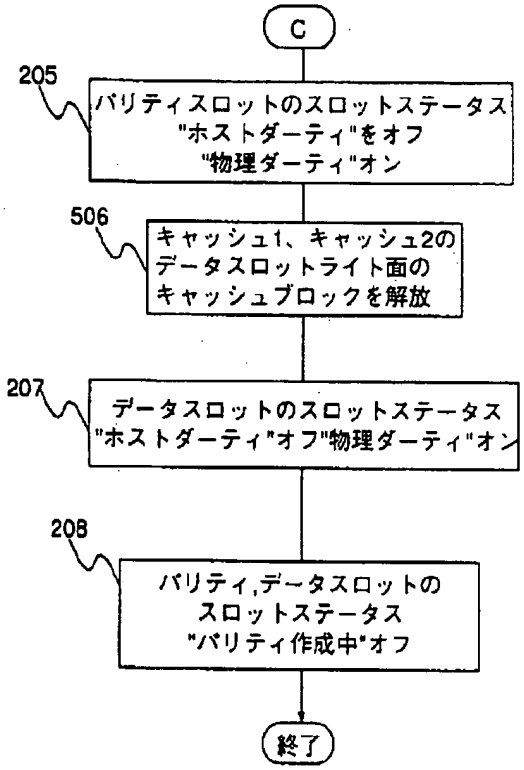
【図12】

図12



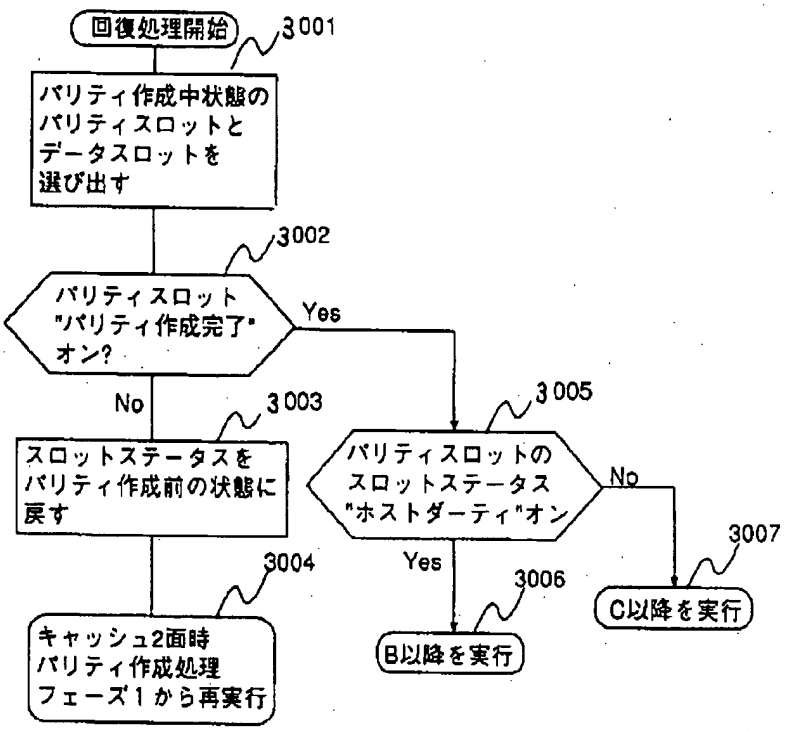
【図13】

図13



【図15】

図15



フロントページの続き

(72)発明者　佐藤　孝夫
神奈川県小田原市国府津2880番地　株式会
社日立製作所ストレージシステム事業部内